Title
Adequacy of SEIR models when epidemics have spatial structure: Ebola in Sierra Leone.

Journal
Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 374(1775)

ISSN
0962-8436

Authors
Getz, Wayne M
Salter, Richard
Mgbara, Whitney

Publication Date
2019-06-01

DOI
10.1098/rstb.2018.0282

Peer reviewed

PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

Research

Cite this article: Getz WM, Salter R, Mgbara W. 2019 Adequacy of SEIR models when epidemics have spatial structure: Ebola in Sierra Leone. *Phil. Trans. R. Soc. B* **374**: 20180282.
http://dx.doi.org/10.1098/rstb.2018.0282

Accepted: 26 February 2019

One contribution of 15 to a theme issue 'Modelling infectious disease outbreaks in humans, animals and plants: approaches and important themes'.

Subject Areas:
computational biology, health and disease and epidemiology

Keywords:
SIR models, R_0, latent period, appropriate complexity models, effective-population-at-risk, time-dependent force-of-infection

Author for correspondence:
Wayne M. Getz
e-mail: wgetz@berkeley.edu

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.4444244.

THE ROYAL SOCIETY PUBLISHING

Adequacy of SEIR models when epidemics have spatial structure: Ebola in Sierra Leone

Wayne M. Getz[1,2], Richard Salter[3] and Whitney Mgbara[1]

[1]Department Environmental Science, Policy and Management, University of California, Berkeley, CA 94708-3112, USA
[2]School of Mathematical Sciences, University of KwaZulu-Natal, Durban, South Africa
[3]Numerus, 850 Iron Point Road, Folsom, CA 95630, USA

WMG, 0000-0001-8784-9354

Dynamic SEIR (Susceptible, Exposed, Infectious, Removed) compartmental models provide a tool for predicting the size and duration of both unfettered and managed outbreaks—the latter in the context of interventions such as case detection, patient isolation, vaccination and treatment. The reliability of this tool depends on the validity of key assumptions that include homogeneity of individuals and spatio-temporal homogeneity. Although the SEIR compartmental framework can easily be extended to include demographic (e.g. age) and additional disease (e.g. healthcare workers) classes, dependence of transmission rates on time, and metapopulation structure, fitting such extended models is hampered by both a proliferation of free parameters and insufficient or inappropriate data. This raises the question of how effective a tool the basic SEIR framework may actually be. We go some way here to answering this question in the context of the 2014–2015 outbreak of Ebola in West Africa by comparing fits of an SEIR time-dependent transmission model to both country- and district-level weekly incidence data. Our novel approach in estimating the effective-size-of-the-populations-at-risk (N_{eff}) and initial number of exposed individuals (E_0) at both district and country levels, as well as the transmission function parameters, including a time-to-halving-the-force-of-infection ($t_{f/2}$) parameter, provides new insights into this Ebola outbreak. It reveals that the estimate $R_0 \approx 1.7$ from country-level data appears to seriously underestimate $R_0 \approx 3.3 - 4.3$ obtained from more spatially homogeneous district-level data. Country-level data also overestimate $t_{f/2} \approx 22$ weeks, compared with 8–10 weeks from district-level data. Additionally, estimates for the duration of individual infectiousness is around two weeks from spatially inhomogeneous country-level data compared with 2.4–4.5 weeks from spatially more homogeneous district-level data, which estimates are rather high compared with most values reported in the literature.

This article is part of the theme issue 'Modelling infectious disease outbreaks in humans, animals and plants: approaches and important themes'. This issue is linked with the subsequent theme issue 'Modelling infectious disease outbreaks in humans, animals and plants: epidemic forecasting and control'.

1. Introduction

Systems of ordinary differential equations in the disease-class variables S (susceptibles), E (exposed), I (infected) and R (removed) have been the mainstay of modelling the dynamics of disease outbreaks for the past 100 years [1,2]. These models have also been discretized with respect to time (systems of discrete-time difference equations) and, in stochastic settings, with respect to state (i.e. integer numbers of individuals in disease classes; for a recent review, see [3]). Although SEIR models assume homogeneity with respect to both disease class and spatial structure, these models have been greatly elaborated to take into account age [4], sex [5], the genetic structure of hosts, and pathogens [6,7]. They have also been

extended to include notions of distributed-delay of individuals within disease classes over time [8], while integral projection methods are used to incorporate continuous traits that may be relevant to disease dynamics [9,10].

Spatial structure has been incorporated into epidemiological models using metapopulation formulations [11–16]. All these extensions lead to the proliferation of model parameters, however, and thus require orders of magnitude more data to fit the model than does the basic three-parameter SEIR disease process formulation with its latent period (Π^E: unit time), infectious period (Π^I: unit time) and *per capita* susceptible transmission rate parameters (β: units vary, depending on how the transmission function is assumed to depend on I and N—see [17,18] for further discussion). If demography is added—as formulated below in terms of recruitment (λ: numbers per unit time) and natural (μ: *per capita* per unit time) and disease-induced mortality (α: *per capita* per unit time)—at least six parameters must be estimated. Also, spatial structure can be accounted for using individual-based models, in which individuals are tagged with their current location on a landscape. However, minimally such models require a movement rate parameter [19] and, perhaps, additional parameters to codify the rules that individuals use to move and encounter other individuals on the landscape [20].

Individual-based (or so-called agent-based) models have the advantage of being able to incorporate the histories of individuals with respect to 'time since a particular disease class transmission event occurred' or the 'genealogy of infection chains' [21–24], with implications for the epidemiological field of phylodynamics [25]. Again, for a comprehensive description of such models, additional parameters are needed that can account for mutation rates and other relevant genetic processes.

To avoid the proliferation of parameters in models, mathematical descriptions of biological populations, by necessity, focus on one or a couple of processes. The challenge is to identify only those processes that are relevant to answering the question at hand about a system that is otherwise a complex concatenation of hierarchically interacting layers of objects and processes [26]. The central question in building models is whether or not they are adequate to address the questions at hand. Adequate models require that we include an appropriate level of structural complexity with regard to the questions being asked and the extent of the empirical data that can support the models [26]. A need for data, of course, is obviated in theoretical studies that explore questions of relevance to classes of systems, such as the properties of epidemics occurring on particular types of contact networks [27–31].

In this paper, we evaluate the adequacy of using SEIVD models—a variant of SEIR models where R is divided into recovered (V: 'naturally vaccinated', i.e. immune) and dead (D)—to forecast and manage epidemiological outbreaks in the context of how robust estimation of epidemic process parameters are, if we ignore obvious inhomogeneities in incidence data. We do so in the context of the Ebola viral disease (EVD) outbreak in West Africa in 2014–2015 [13,21,22,32–34]. Specifically, we compare the fit of a discrete-time SEIVD model with time-varying transmission parameter $\beta(t)$ [3] to weekly incidence data among the (i) combined Freetown and Waterloo (F&W) districts in Sierra Leone, (ii) the Port Loko (PL) district in Sierra Leone and (iii) Sierra Leone (SL) as a whole. In some cases, we allow the initial number of exposed individuals $E(0)$ to be a fitted parameter, as a way of dealing with the fact that we do not know when the first infection actually occurred within a particular district. Further, we allow for the fact that the actual size of the susceptible group underlying a particular set of incidence data is not known by fitting an effective-population size N_{eff}, a term borrowed from genetics, but used here in a different context. Specifically, N_{eff} answers the question: what would the best estimate of the size of the underlying population be if the epidemic were taking place in a homogeneous population (i.e. homogeneous individual hosts from, *inter alia*, a spatial, behavioural, physiological and immunological point of view; as well as a single homogeneous strain of pathogen)?

2. Model construction

The model we fit is a discrete-time deterministic version of the SEIVD model presented in Getz *et al.* [3]. SEIR epidemic models are typically formulated as systems of ordinary differential equations (ODEs; see [2]) in the variables S (susceptible), E (infected but not yet infectious), I (infectious) and R (removed: further divided into immune V and dead D). Discrete-time equivalents of these models are particularly useful when fitting models to data collected at a fixed rate (e.g. daily, weekly or monthly rates) or evaluating interventions (treatment regimens, isolation of patients) that are themselves discrete (e.g. daily rates) [3,18,35,36]. Further, these discrete formulations are more easily extended than continuous time formulations to stochastic settings, where the latter require implementation of event-driven formulations [3,8,18,37–39].

In the development of our model, as mentioned above, we break R into V (recovered with temporary immunity) and D (dead) [3,18] (electronic supplementary material, figure S1). Also, we use the roman fonts X = S, E, I, V and D to name classes and the italic fonts $X = S,\ E,\ I,\ V$ and D to represent the values of the corresponding variables. An additional value that appears in most epidemic models is the total population alive at time t: i.e. $N(t) = S(t) + E(t) + I(t) + V(t)$.

Our model has four epidemiological flow rates: γ^X, X = S, E, I or V. The latter three are the reciprocals of the average time an individual spends in class X: i.e. the latent period $\Pi^E = 1/\gamma^E$ is the mean time spent in state E, the infectious period $\Pi^I = 1/\gamma^I$ is the mean time spent in state I, and the refractory immune period $\Pi^V = 1/\gamma^V$ is the mean time spent in state V (where the designation 'mean' is strictly only meaningful in a stochastic version of this model—see [3]). For Ebola, lifelong immunity implies that the rate γ^V is 0. The *per capita*-S transmission rate γ^S, however, is not a parameter, but a transmission rate function that either scales with I (mass-action transmission), with I/N (frequency-dependent transmission function—see [40]) or is a more general function of I and N [17,41]. As epidemics unfold, γ^S may explicitly depend on time t if individuals adjust their behaviour to reduce contact with infected individuals (avoiding crowds) or the probability of disease transmission when contact is made (e.g. healthcare workers wearing masks and using gloves).

Though the dependence of γ^S on t is most often assumed to be exponential decay (e.g. [32]), we prefer an expression of time that has slope 0 at the origin because of its

obvious greater realism [42]: e.g. the generalized three-parameter ($\beta > 0$, $t_{f/2} > 0$ and $\varepsilon > 1$) reverse S-shaped hyperbolic expression [3,21]

$$\gamma^S(I, N, t) = \frac{\beta I}{(1 + (t/t_{f/2})^{\varepsilon})N}. \tag{2.1}$$

Specifically, the value of $\gamma^S(I, N, t)$ as a function of time t falls off in value for a give ratio I/N from $\gamma^S(I, N, 0) = \beta I/N$ to $\lim_{t\to\infty}\gamma^S(I, N, t) = 0$, passing through the half-way mark $\beta I/(2N)$ at the critical time $t_{f/2}$, with a steepness (also known as the abruptness parameter—see [42]) that increases with the value of ε (electronic supplementary material, figure S2). Of course, in practice, additional dependence on time enters through the fact that I and N are themselves functions of t.

Demography is added to the model by including natural and disease-induced *per capita* death rates parameters, μ and α respectively, as illustrated in electronic supplementary material, figure S1, while recruitment is included through the parameter Λ. This recruitment could be generated through births or actual recruitment from younger cohorts (as in models that consider sexually transmitted diseases) or from migration (as in metapopulation models). The equations for this discrete SEIVD model are provided in electronic supplementary material, equations S.1–S.8, and also in Getz *et al.* [3], where the expression for R_0—the expected number of susceptibles to be infected by the index case—is provided by the expression

$$R_0 = \frac{\beta\gamma^E}{(\gamma^E + \mu)(\gamma^I + \mu + \alpha)}. \tag{2.2}$$

(a) Fitting our model to data

We introduce several novelties in our approach to fitting our SEIVD model to weekly incidence data. First, besides the usual approach of fitting the parameters β (transmission), $\Pi^E = 1/\gamma^E$ (latent period) and $\Pi^I = 1/\gamma^I$ (infectious period), we also fit $t_{f/2}$ (time to halving the force of infection), $N_{\text{eff}} = N(0)$ (the effective size of the population at risk) and $E(0) = E_0$, (the unknown number of exposed individuals at the start of the infection), where the outbreak starts with the first recorded infectious case (and the index case and others if there is more than one infectious case).

For certain types of outbreaks—those where the number of infectious individuals over the course of the epidemic remains small compared with the total population size—the best model fit may be rather insensitive to the selection of a value for $N(0) = N_{\text{eff}}$. The reason is that the transmission rate (as derived by multiplying the *per capita* susceptible transmission rate, expressed in equation (2.1), by S), in situations when $I(t) \ll N(t) \Rightarrow S(t) \approx N(t)$, is approximately given by

$$\begin{aligned}\gamma^S((I(t), N(t), t)S(t) &= \frac{\beta I(t)}{(1 + (t/t_{f/2})^{\varepsilon})N(t)} S(t)\\ &\approx \frac{\beta I(t)}{1 + (t/t_{f/2})^{\varepsilon}}\end{aligned} \tag{2.3}$$

which is independent of $N(t)$ and hence $N_{\text{eff}} = N(0)$. Only in outbreaks where $I(t)$ rises above some significant fraction of $N(t)$ (say, $I(t) > 0.05N(t)$) will our approach to estimating the size of N_{eff} begin to yield reliable insights into the size of the population at risk.

Parameter estimation is implemented on the Numerus Model Builder platform [3] using a Nelder–Mead optimization procedure. We fitted our model to the data using a maximum-likelihood procedure described in Getz *et al.* [3]. Initial runs revealed the presence of several local optima. Thus, we selected the best solution in each case (we had three sets of data and allowed from three to seven parameters to vary, as described in the next section) for 100 to 400 runs starting from random initial selections of parameter values (the Nelder–Mead requires two initial estimates to get started).

(b) Ebola viral disease in Sierra Leone

At a coarse spatial level, Sierra Leone can be divided into four geographical regions, one of which is centred around the only major city in Sierra Leone, Freetown, which has a million plus individuals. The other three regions each contain at most two towns with between 0.1 and 0.2 million individuals (electronic supplementary material, figure S4). Without more refined spatial population data than this, as well as some information on how individuals move among the major population centres (i.e. towns with ten thousand or more individuals apiece) or from adjoining rural areas to these centres, a reliable analysis of the spatial aspects of an Ebola outbreak in Sierra Leone cannot be undertaken. In addition, an integrated spatial analysis would require information on individuals moving from the neighbouring states of Liberia and Guinea to account for the across-border transmission of the Ebola virus.

Although the 2014–2015 West African EVD outbreak has received considerable attention from the modelling and simulation community [13,21,32,33,43], many questions relating to the impact of spatio-temporal inhomogeneity on the epidemic parameters remain unanswered. Here we use a deterministic, discrete-time, SEIVD model (electronic supplementary material, appendix; also see [3]) to separately fit different Sierra Leone districts, as well as country-level weekly incidence data to explore the question of how the epidemic varied among regions. The regional weekly incidence data we fitted are reported in Fang *et al.* [33], and for purposes of illustration plotted in electronic supplementary material, figure S3. From these plots, we see that outbreaks peaked at different times in different districts of Sierra Leone. The outbreak started in the eastern-most town and district of Kailahun (electronic supplementary material, figure S4) and peaked last in the western-most region containing the capital Freetown and the neighbouring city of Waterloo (Western district). A relatively large outbreak peaked in Port Loko half-way through the 52-week epidemic. The outbreaks in Freetown–Waterloo (F&W) and Port Loko (PL) are much more typical of those seen in homogeneous populations than is the wider incidence peak observed in Sierra Leone (SL) as a whole: wide peaks suggest spatial heterogeneity in the data or a different set of epidemiological parameters—as is demonstrated by our analysis below. Thus, we decided to compare how well our SEIVD model could be fitted to F&W, PL and SL data, ignoring the smaller outbreaks in other districts as generally too noisy to provide good fits to a deterministic model. For a discussion on this latter point in the context of the Liberian EVD outbreak, see Getz *et al.* [36], which contains empirical data showing that small Liberian towns and villages were susceptible to repeated small outbreaks. These latter phenomena are likely

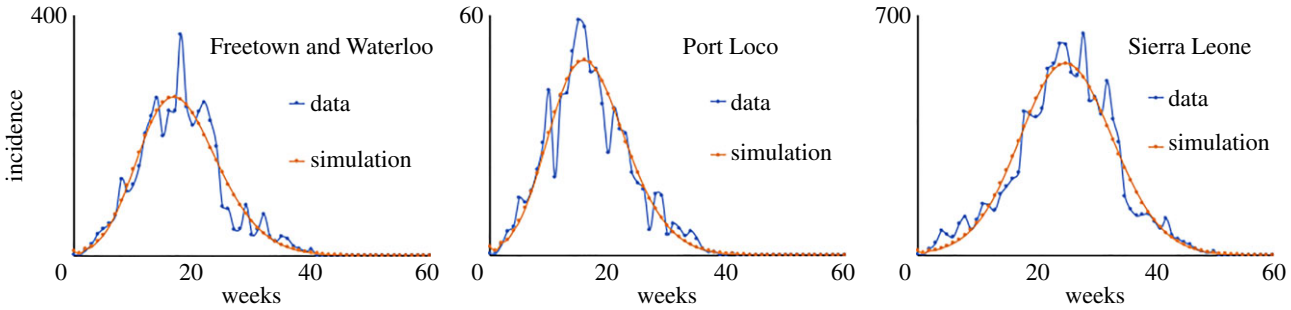

Figure 1. Best three-parameter fits of a homogeneous SEIR model to selected relatively spatially cohesive district data (F&W, Western Area; PL, Northern Area) and clearly spatially structured country-level data (SL) plotted against weeks since the beginning of the specific local (rather than country-wide) outbreak (see table 1 for parameter, maximum-likelihood and AIC values).

Table 1. AIC [44,45] fits of models to incidence data treated for purposes of the fits as closed homogeneous districts within the country of Sierra Leone. Italicized values denote fixed parameter values, while the remaining are varied using an NMB implementation of the Nelder Mead algorithm (NMA, [46]) to maximize the log-likelihood (LL), and hence the AIC, of the fit. For the three and four free parameter cases, the fixed parameter values are set close to those estimated in the seven parameter case to allow us to evaluate the robustness of our fitting procedures in estimating the remaining free parameter values. NMA starting values are selected at random from the bottom and top half of the indicated parameter ranges. The reported maximum is the best of 100 runs in each case (except for the seven parameter country case where 400 runs were compared), using lower and upper starting values from each parameter, respectively, randomly selected from the lower and upper halves of the indicated search range. F&W is the combined Freetown and Waterloo districts; PL is the Port Loko district and SL is the whole of Sierra Leone. See electronic supplementary material, figure S4 for sources on setting local population size.

data (units)	# free params.	N_{eff} (millions)	E_0 (#)	Π^E (weeks)	Π^I (weeks)	$\frac{\beta}{1+(t/t_{f/2})^{\varepsilon}}$	R_0[a]	LL	AIC *best
F&W	range	[0.05, 1.5]	[1, 20]	[0.1,2]	[2,6]	$\beta \in [1, 3]$, $t_{f/2} \in [3, 11]$, $\varepsilon \in [1, 3]$			
	7	0.645	10.4	1.07	4.23	$\frac{2.10}{1+(t/8.23)^{2.04}}$	4.8	−284.8	584
	4	0.581	12.8	*1*	*4*	$\frac{1.94}{1+(t/8.59)^{2}}$	4.3	−284.8	578
	3	*1.1*	11.7	*1*	*4*	$\frac{1.96}{1+(t/8.58)^{2}}$	4.3	−284.6	575*
PL	range	[0.05, 0.8]	[1, 20]	[0.1, 2]	[2, 6]	$\beta \in [1, 3]$, $t_{f/2} \in [3, 11]$, $\varepsilon \in [1, 3]$			
	7	0.227	7.5	0.71	2.61	$\frac{2.00}{1+(t/7.62)^{1.69}}$	3.4	−170.6	355
	4	0.331	11.2	*1*	*3*	$\frac{1.67}{1+(t/9.60)^{2}}$	3.1	−173.6	354*
	3	*0.55*	9.0	*1*	*3*	$\frac{1.79}{1+(t/9.20)^{2}}$	3.3	−174.4	354*
SL	range	[4, 12]	[5, 25]	[0.1, 2]	[1, 4]	$\beta \in [0.5, 2]$, $t_{f/2} \in [10, 25]$, $\varepsilon \in [1, 3]$			
	7	10.903	18.4	1.06	1.78	$\frac{1.29}{1+(t/22.07)^{2.50}}$	1.7	−471.6	957
	4	11.569	20.0	*1*	*2*	$\frac{1.21}{1+(t/22.27)^{2.5}}$	1.7	−465.0	938
	3	*7*	20.0	*1*	*2*	$\frac{1.21}{1+(t/22.28)^{2.5}}$	1.7	−465.0	936*

[a]Using equation (2.2) for the case $\mu = 0.001$ and $\alpha = 0.2$.

due to individuals moving among population concentrations reinvigorating outbreaks in areas where epidemics may have been waning.

The results of our SEIVD model fits are illustrated in figure 1 and reported in table 1. Due to the lack of appropriate movement data on the sweep of the Sierra Leone 2014–2015 EVD across the country from east to west [33], we did not try to fit a metapopulation version of our SEIR model, using the approach discussed in Getz *et al*. [3] to extending a continuous time SEIVD model to a metapopulation setting.

3. Discussion

The results we obtain indicate very clearly that when an epidemic has a clear spatial structure—for example, indicated

outbreaks starting earlier in one region and taking off later in another—then incidence has a broader distribution (e.g. around 50 weeks long in SL versus 40 weeks each in F&W and PL, as can be seen in figure 1) resulting in larger estimates for time-to-halving-the-force-of-infection (e.g. $t_{f/2}$ is around nine weeks in the two districts versus 22 weeks for Sierra Leone as a whole). The broadening of the outbreak distribution for a given maximum incidence peak implies that estimates for the transmission parameter are lower (e.g. β is around 2 in the districts versus around 1.25 in SL) which, in turn, influences the estimate of R_0 (equation (2.2)).

Our analysis of district-level data suggests that R_0 is somewhere in the range of 3–5, while analysis of country-level data suggests that $R_0 \approx 1.7$. Estimates of R_0 in Guinea, Sierra Leone and Liberia around four months into epidemics in the latter two countries (i.e. around mid to late August of 2014) [32], yield the respective values 1.52, 2.42 and 1.65, using an exponential decay model for β (electronic supplementary material, figure S2). The Guinea and Liberia estimates are in line with what we get here for Sierra Leone as a whole, although the estimate for Sierra Leone is somewhat higher than what we obtain here (2.42 versus 1.7). The estimate we obtain here for Sierra Leone is very similar to the estimates we obtained using an individual-based stochastic process model, which produced a value of $R_0 \approx 1.6$ [21]. Values for R_0 around 4 are more in line with the values we recently obtained in fitting an SEIVD distributed-delay (Erlang or boxcar—see [8]) model to the Liberian incidence data, where $R_0 \approx 4.5$ was obtained.

Since transmission is a function of both pathogen and host biology, as well as environmental factors related to host location and behaviour, we should expect R_0 to vary across communities with different cultural practices and economic circumstances. Our results suggest, however, in using a spatially homogeneous model to fit data from the spatially structured outbreak, that R_0 is greatly underestimated, with consequences for managing epidemics including assessment of vaccination coverage levels needed to obtain 'herd immunity' conditions within a vulnerable population [2]. On the other hand, if R_0 levels are relatively high, then the obvious question is why have so many Ebola outbreaks been relatively small or exhibited the kind of stuttering behaviour that is associated with values of R_0 not much larger than 1 [47]. The answer may lie in the fact that many of the Ebola outbreaks in various parts of central and east Africa, from the 1970s until recently, have occurred in low density, relatively isolated locations. Here epidemic fade-out occurs because of small localized population sizes [48], unlike in towns and cities where populations of several hundred thousand or more are able to fuel epidemics involving hundreds or thousands of cases.

Our analysis also reveals that, although the fitting procedures are insensitive to assumptions regarding the initial size (N_{eff}) of the population at risk, estimates obtained are still within a factor of 2 of the listed population densities within the districts themselves. This insensitivity, as discussed in the context of equation (2.3), is not unexpected; and the results we obtained suggest that for the transmission process that has a greater density-dependent component, as is likely for epizootic compared with epidemic processes [41], fitting this parameter could prove to be useful in data-rich situations.

Our analysis indicates that beyond individual variation in transmission (e.g. superspreaders [49]) or resistance to infection, accurate assessment of R_0 requires that we are cognizant of the degree to which the data come from a well-mixed, spatially homogeneous population. To get an assessment of spatial heterogeneity requires data on variation in population density or the movement of individuals. The model we formulated can be extended to include additional disease states (e.g. symptomatic but uninfectious [50]). In the context of Ebola, this includes transmission from individuals being prepared for burial [51].

The model we present here can be extended to a metapopulation setting, as has been done in the context of continuous SEIR models [3]. The drawback, however, is that appropriate data are often not available to adequately support these additional structures. Many wildlife populations have a metapopulation structure [52], and diseases in such populations, including rinderpest in artiodactyls and canine distemper in carnivores [53], has important species conservation implications [54]. Metapopulation models have also been used to study the spread of disease in plant populations, with the anther-smut disease in *Lychnis alpina* being a case in point [55]. Metapopulation structures naturally arise through husbandry practices in domestic animals, likely involving many more subpopulations than in the wild: each subpopulation corresponds to a common ranching area, a group of contiguous farms, large individual farms themselves or individual poultry rearing facilities. For example, individual farms were treated as separate subpopulations in modelling the outbreak of foot-and-mouth disease in England in 2001 [27]. The application of metapopulation models to disease outbreaks in domesticated animals is facilitated by the fact that transfers of individuals among subpopulations may be well documented, and the limitation of such transfers used to control outbreaks [56]. Crop production mosaics may also take on a metapopulation structure in the context of epidemic outbreaks, where the disease is vectored among different parts of the mosaic by insects or other types of vectors, including wind and water [57]. Gilligan [58], for example, proposed that each field should be considered as a 'natural unit for infection'.

Finally, our analysis suggests that fitting models with more than a half-dozen parameters is exceptionally challenging because of both the 'curse of dimensionality' [59] and the impacts of multi-collinearity. Fitting such models would likely require that different parts of the model be fitted independently to different datasets. For example, a movement component, when employed, could first be fitted to a movement model, or in the case of EVD the infectious period could be set using the consensus level of around 9.5 days [60] (our high estimate of this quantity might be explained by our neglect of transmission during the preparation of bodies for burial [51]). Then epidemiological parameters could be obtained by fitting an SEIVD model to subsets of incidence data for which the assumption of well-mixed subpopulations is reasonable. Once these components of a larger model have been fitted, then the larger model itself can be used to make a logistical decision on how best to manage epidemics that threaten large regions around the globe.

Data accessibility. This article has no additional data.

Competing interests. We declare we have no competing interests.

Funding. W.M.G. was supported in part by the A. Starker Leopold Chair of Wildlife Ecology at UC Berkeley. R.S. was supported in part by NSF/CPATH-2 CNS0939153. W.M.G. was supported by a UC Berkeley Chancellor's Graduate Student Fellowship.

References

1. Smith DL, Battle KE, Hay SI, Barker CM, Scott TW, McKenzie FE, Chitnis CE. 2012 Ross, Macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens. *PLoS Pathog.* **8**, e1002588. (doi:10.1371/journal.ppat.1002588)
2. Hethcote HW. 2000 The mathematics of infectious diseases. *SIAM Rev.* **42**, 599–653. (doi:10.1137/S0036144500371907)
3. Getz WM, Salter R, Muellerklein O, Yoon HS, Tallam K. 2018 Modeling epidemics: a primer and Numerus Model Builder implementation. *Epidemics* **25**, 9–19. (doi:10.1016/j.epidem.2018.06.001)
4. Castillo-Chavez C, Hethcote HW, Andreasen V, Levin SA, Liu WM. 1989 Epidemiological models with age structure, proportionate mixing, and cross-immunity. *J. Math. Biol.* **27**, 233–258 (doi:10.1007/BF00275810)
5. Leclerc PM, Matthews AP, Garenne ML. 2009 Fitting the HIV epidemic in Zambia: a two-sex micro-simulation model. *PLoS ONE* **4**, e5439. (doi:10.1371/journal.pone.0005439)
6. Koelle K, Cobey S, Grenfell B, Pascual M. 2006 Epochal evolution shapes the phylodynamics of interpandemic influenza A (H_3N_2) in humans. *Science* **314**, 1898–1903. (doi:10.1126/science.1132745)
7. Gilchrist MA, Sasaki A. 2002 Modeling host–parasite coevolution: a nested approach based on mechanistic models. *J. Theor. Biol.* **218**, 289–308. (doi:10.1006/jtbi.2002.3076)
8. Getz WM, Dougherty ER. 2018 Discrete stochastic analogs of Erlang epidemic models. *J. Biol. Dyn.* **12**, 16–38. (doi:10.1080/17513758.2017.1401677)
9. Rees M, Ellner SP. 2009 Integral projection models for populations in temporally varying environments. *Ecol. Monogr.* **79**, 575–594. (doi:10.1890/08-1474.1)
10. Metcalf CJE, Graham AL, Martinez-Bakker M, Childs DZ. 2016 Opportunities and challenges of Integral Projection Models for modelling host-parasite dynamics. *J. Anim. Ecol.* **85**, 343–355. (doi:10.1111/1365-2656.12456)
11. Lloyd AL, Jansen VA. 2004 Spatiotemporal dynamics of epidemics: synchrony in metapopulation models. *Math. Biosci.* **188**, 1–16. (doi:10.1016/j.mbs.2003.09.003)
12. Watts DJ, Muhamad R, Medina DC, Dodds PS. 2005 Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proc. Natl Acad. Sci. USA* **102**, 11 157–11 162. (doi:10.1073/pnas.0501226102)
13. Kramer AM, Tomlin Pulliam J, Alexander LW, Park AW, Rohani P, Drake JM. 2016 Spatial spread of the West Africa Ebola epidemic. *Open Sci.* **3**, 160294. (doi:10.1098/rsos.160294)
14. Keeling MJ. 1999 The effects of local spatial structure on epidemiological invasions. *Proc. R. Soc. Lond. B* **266**, 859–867. (doi:10.1098/rspb.1999.0716)
15. Balcan D, Gonçalves B, Hu H, Ramasco JJ, Colizza V, Vespignani A. 2010 Modeling the spatial spread of infectious diseases: the GLobal Epidemic and Mobility computational model. *J. Comput. Sci.* **1**, 132–145. (doi:10.1016/j.jocs.2010.07.002)
16. Ajelli M, Gonçalves B, Balcan D, Colizza V, Hu H, Ramasco JJ, Merler S, Vespignani A. 2010 Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models. *BMC Infect. Dis.* **10**, 190. (doi:10.1186/1471-2334-10-190)
17. McCallum H, Barlow N, Hone J. 2001 How should pathogen transmission be modelled? *Trends Ecol. Evolut.* **16**, 295–300. (doi:10.1016/S0169-5347(01)02144-9)
18. Getz WM, Lloyd-Smith JO. 2006 Basic methods for modeling the invasion and spread of contagious diseases. In *Disease evolution: models, concepts, and data analyses* (eds Z Feng, U Dieckmann, SA Levin), pp. 87–112. DIMACS vol. 71. Providence, RI: American Mathematical Society.
19. Cross PC, Lloyd-Smith JO, Johnson PL, Getz WM. 2005 Duelling timescales of host movement and disease recovery determine invasion of disease in structured populations. *Ecol. Lett.* **8**, 587–595. (doi:10.1111/ele.2008.8.issue-6)
20. Getz WM, Salter R, Lyons AJ, Sippl-Swezey N. 2015 Panmictic and clonal evolution on a single patchy resource produces polymorphic foraging guilds. *PLoS ONE* **10**, e0133732. (doi:10.1371/journal.pone.0133732)
21. Getz WM *et al.* 2015 Tactics and strategies for managing Ebola outbreaks and the salience of immunization. *Comput. Math. Methods Med.* **2015**, 736507. (doi:10.1155/2015/736507)
22. Scarpino SV *et al.* 2014 Epidemiological and viral genomic sequence analysis of the 2014 Ebola outbreak reveals clustered transmission. *Clin. Infect. Dis.* **60**, 1079–1082. (doi:10.1093/cid/ciu1131)
23. Kiskowski MA. 2014 A three-scale network model for the early growth dynamics of 2014 West Africa Ebola epidemic. *PLoS Curr.* **6**. (doi:10.1371/currents.outbreaks.c6efe8274dc55274f05cbcb62bbe6070)
24. Kühnert D, Stadler T, Vaughan TG, Drummond AJ. 2014 Simultaneous reconstruction of evolutionary history and epidemiological dynamics from viral sequences with the birth-death SIR model. *J. R. Soc. Interface* **11**, 20131106. (doi:10.1098/rsif.2013.1106)
25. Frost SDW, Pybus OG, Gog JR, Viboud C, Bonhoeffer S, Bedford T. 2015 Eight challenges in phylodynamic inference. *Epidemics* **10**, 88–92. (doi:10.1016/j.epidem.2014.09.001)
26. Getz WM *et al.* 2018 Making ecological models adequate. *Ecol. Lett.* **21**, 153–166. (doi:10.1111/ele.12893)
27. Keeling MJ, Eames KT. 2005 Networks and epidemic models. *J. R. Soc. Interface* **2**, 295–307. (doi:10.1098/rsif.2005.0051)
28. Eames K, Bansal S, Frost S, Riley S. 2015 Six challenges in measuring contact networks for use in modelling. *Epidemics* **10**, 72–77. (doi:10.1016/j.epidem.2014.08.006)
29. Castellano C, Pastor-Satorras R. 2010 Thresholds for epidemic spreading in networks. *Phys. Rev. Lett.* **105**, 218701. (doi:10.1103/PhysRevLett.105.218701)
30. Youssef M, Scoglio C. 2011 An individual-based approach to SIR epidemics in contact networks. *J. Theor. Biol.* **283**, 136–144. (doi:10.1016/j.jtbi.2011.05.029)
31. Miller JC, Volz EM. 2013 Incorporating disease and population structure into models of SIR disease in contact networks. *PLoS ONE* **8**, e69162. (doi:10.1371/journal.pone.0069162)
32. Althaus CL. 2014 Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa. *PLoS Curr.* **6**. (doi:10.1371/currents.outbreaks.91afb5e0f279e7f29e7056095255b288)
33. Fang L-Q *et al.* 2016 Transmission dynamics of Ebola virus disease and intervention effectiveness in Sierra Leone. *Proc. Natl Acad. Sci. USA* **113**, 4488–4493. (doi:10.1073/pnas.1518587113)
34. Backer JA, Wallinga J. 2016 Spatiotemporal analysis of the 2014 Ebola epidemic in West Africa. *PLoS Comput. Biol.* **12**, e1005210. (doi:10.1371/journal.pcbi.1005210)
35. Lloyd-Smith JO, Galvani AP, Getz WM. 2003 Curtailing transmission of severe acute respiratory syndrome within a community and its hospital. *Proc. R. Soc. Lond. B* **270**, 1979–1989. (doi:10.1098/rspb.2003.2481)
36. Getz WM, Lloyd-Smith JO, Cross PC, Bar-David S, Johnson PL, Porco TC, Sánchez MS. 2006 Modeling the invasion and spread of contagious disease in heterogeneous populations. In *Disease evolution: models, concepts, and data analyses* (eds Z Feng, U Dieckmann, SA Levin), pp. 113–144. DIMACS vol. 71. Providence, RI: American Mathematical Society.
37. Gillespie DT. 1976 A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **22**, 403–434. (doi:10.1016/0021-9991(76)90041-3)
38. Allen LJ. 2017 A primer on stochastic epidemic models: formulation, numerical simulation, and analysis. *Infect. Dis. Model.* **225**, 128–142.
39. Britton T. 2010 Stochastic epidemic models: a survey. *Math. Biosci.* **225**, 24–35. (doi:10.1016/j.mbs.2010.01.006)
40. Getz WM, Pickering J. 1983 Epidemic models: thresholds and population regulation. *Am. Nat.* **121**, 892–898. (doi:10.1086/284112)
41. Lloyd-Smith JO *et al.* 2005 Should we expect population thresholds for wildlife disease? *Trends Ecol. Evolut.* **20**, 511–519. (doi:10.1016/j.tree.2005.07.004)
42. Getz WM. 1996 A hypothesis regarding the abruptness of density dependence and the growth rate of populations. *Ecology* **77**, 2014–2026. (doi:10.2307/2265697)
43. Rivers CM, Lofgren ET, Marathe M, Eubank S, Lewis BL. 2014 Modeling the impact of interventions on an epidemic of Ebola in Sierra Leone and Liberia.

PLoS Curr. **6**. (doi:10.1371/currents.outbreaks.4d41fe5d6c05e9df30ddce33c66d084c)

44. Burnham KP, Anderson DR. 2004 Multimodel inference: understanding AIC and BIC in model selection. *Sociol. Methods Res.* **33**, 261–304. (doi:10.1177/0049124104268644)
45. Burnham KP, Anderson DR. 2003 *Model selection and multimodel inference: a practical information-theoretic approach*. Berlin, Germany: Springer Science & Business Media.
46. Singer S, Nelder J. 2009 Nelder-Mead algorithm. *Scholarpedia* **4**, 2928. (doi:10.4249/scholarpedia.2928)
47. Lloyd-Smith JO, George D, Pepin KM, Pitzer VE, Pulliam JRC, Dobson AP, Hudson PJ, Grenfell BT. 2009 Epidemic dynamics at the human-animal interface. *Science* **326**, 1362–1367. (doi:10.1126/science.1177345)
48. Krauer F, Gsteiger S, Low N, Hansen CH, Althaus CL. 2016 Heterogeneity in district-level transmission of Ebola virus disease during the 2013–2015 epidemic in West Africa. *PLoS Negl. Trop. Dis.* **10**, e0004867. (doi:10.1371/journal.pntd.0004867)
49. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. 2005 Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359. (doi:10.1038/nature04153)
50. Thompson RN, Hart WS. 2018 Effect of confusing symptoms and infectiousness on forecasting and control of Ebola outbreaks. *Clin. Infect. Dis.* **67**, 1472–1474. (doi:10.1093/cid/ciy248)
51. Tiffany A *et al.* 2017 Estimating the number of secondary Ebola cases resulting from an unsafe burial and risk factors for transmission during the West Africa Ebola epidemic. *PLoS Negl. Trop. Dis.* **11**, e0005491. (doi:10.1371/journal.pntd.0005491)
52. McCullough DR 1996 *Metapopulations and wildlife conservation*. Washington, DC: Island Press.
53. Deem SL, Spelman LH, Yates RA, Montali RJ. 2000 Canine distemper in terrestrial carnivores: a review. *J. Zoo Wildl. Med.* **31**, 441–451. (doi:10.1638/1042-7260(2000)031[0441:CDITCA]2.0.CO;2)
54. Hess G. 1996 Disease in metapopulation models: implications for conservation. *Ecology* **77**, 1617–1632. (doi:10.2307/2265556)
55. Carlsson-Granér U, Thrall PH. 2002 The spatial distribution of plant populations, disease dynamics and evolution of resistance. *Oikos* **97**, 97–110. (doi:10.1034/j.1600-0706.2002.970110.x)
56. Fèvre EM, Bronsvoort BMDC, Hamilton KA, Cleaveland S. 2006 Animal movements and the spread of infectious diseases. *Trends Microbiol.* **14**, 125–131. (doi:10.1016/j.tim.2006.01.004)
57. Cunniffe NJ, Koskella B, Metcalf CJE, Parnell S, Gottwald TR, Gilligan CA. 2015 Thirteen challenges in modelling plant diseases. *Epidemics* **10**, 6–10. (doi:10.1016/j.epidem.2014.06.002)
58. Gilligan CA. 2007 Sustainable agriculture and plant diseases: an epidemiological perspective. *Phil. Trans. R. Soc. B* **363**, 741–759. (doi:10.1098/rstb.2007.2181)
59. Keogh E, Mueen A2011 *Encyclopedia of machine learning*, pp. 257–258. Berlin, Germany: Springer.
60. Velásquez GE, Aibana O, Ling EJ, Diakite I, Mooring EQ, Murray MB. 2015 Time from infection to disease and infectiousness for Ebola virus disease, a systematic review. *Clin. Infect. Dis.* **61**, 1135–1140. (doi:10.1093/cid/civ531)